科学大数据管理技术与系统

黎建辉[1*] 李跃鹏[1,2] 王华进[1] 陈明奇[3*]

1 中国科学院计算机网络信息中心 北京 100190
2 中国科学院大学 北京 100049
3 中国科学院 办公厅 北京 100864

摘要 由于现代科学发现越来越依赖于大规模科学数据的分析处理，如何高效管理科学大数据业已成为当下亟待解决的问题。文章分析了科学大数据的应用场景和需求，阐述了科学大数据在规模动态化、流水线管理、统一访问、数据共享（SPUS）4个方面面临的挑战。提出了包括计算和存储管理、数据流水线管理、数据融合查询管理、数据共享管理4个模块的科学大数据管理系统体系结构，并分析了系统中存在的关键技术问题。最后，介绍了国家重点研发计划项目“科学大数据管理系统”的研发进展及其未来的研究方向。

关键词 科学大数据，融合查询，流水线，数据共享，弹性伸缩

DOI 10.16418/j.issn.1000-3045.2018.08.005

Jim Gray[1]提出了科学研究的第四范式——数据密集型科学发现的观点，他认为海量数据是未来驱动科学发现的主要动力之一。2012 年 7 月 4 日，欧洲粒子物理中心（CERN）通过分析过去两年大型强子对撞机（LHC）的实验数据，宣布发现“上帝粒子”；次年，“上帝粒子”预言者获得了诺贝尔物理学奖。激光干涉仪引力波天文台（LIGO）科学合作组织在积累 500 PB 数据、历时 14 年模型和系统改进以后，2016 年 2月11日宣布第一次探测到了引力波的存在，证实了相对论的最后预言；2017 年LIGO 的 3 位重要贡献者获得了诺贝尔物理学奖。环顾当今的重大科学研究装置和项目，如天文领域的大型巡天望远镜（LSST）、高能物理领域的大型强子对撞机（LHC）、生命科学领域的人类基因组计划（HGP）、地球科学领域的灾害风险综合研究计划（IRDR）等，无一不是从大科学装置或观测设备中持续不断采集数据，然后通过数据分析进行科学发现。毫无疑问，如今的科学发现模式已经进入科学大数据驱动的时代。到 2020 年左右，LSST 将全面完工运行，届时 LSST 每 3 天完成 1 次巡天，每天产生 15 TB 数据以用于新星发现、暗物质探测等科研目标；阵列射电望远镜（SKA）每秒将产生 200 GB 原始数据、每秒千万亿次计算、10 倍于现有因特网传输速度，正等待科研人员去

*通讯作者

资助项目：国家重点研发计划（2016YFB1000600），国家自然科学基金项目（91546125）
修改稿收到日期：2018年8月15日

突破和挑战。这些大科学项目对于宇宙起源认识、自然规律发现、科技创新具有重大意义，能否有效管理、处理、利用这些数据，将成为我国在新时代下能否取得国际科技领先地位的关键因素之一。

1 科学大数据应用场景及管理需求

1.1 科学大数据的应用场景及典型特征

科学数据是科研活动的输入、输出和资产，是证实或者证伪科学发现或科学观点事实、证据或者论证推理的基础。它包括数字化观测、科学监测等来自仪器设备或传感器的数据，计算模拟与模型输出的数据，对情景或现象的描述，对行为的观测或定性描述，以及用于管理或者商业目的的统计数据等[2]。目前科学大数据普遍存在于各个领域的科学研究，尤其在天文学、高能物理、微生物学等大科学领域，科学大数据的应用场景尤为明显[3]。

在天文学领域，中法合作伽马暴探测天文卫星SVOM的关键地面设备GWAC的每个相机15 s内会产生32 MB的天区图，并于下一个天区图产生之前完成点源提取、交叉认证等操作，最终在3—5 s内完成100万—10 000万行星表数据的插入，10亿—100亿行星表数据的JOIN运算[4]。

在高能物理领域，欧洲核子物理研究组织构建的大型强子对撞机（LHC）每秒进行6亿次碰撞实验，产生6 PB事例数据，经事例筛选后存储大约1 GB实验数据。目前LHC产生的实验数据已超过200 PB，未来5年LHC产生的数据将会超过1 EB，事例数将达到千万亿级别，需在10 s内完成百万分之一的事例筛选操作[5]。

在微生物学领域，中国科学院微生物研究所世界数据中心（WDCM）对Taxonomy、GenBank、Gene等36个数据源进行实体识别、歧义消除、本体构建等数据处理操作，构建了包含830万个节点、1.3亿条边的知识图谱结构。预计未来5年内，WDCM还将汇聚开放生物资源、文献、序列和疾病等数据，在10 000多个数据源中构建100亿条关联的知识图谱数据，并要求1 s内完成100亿条关联数据的6步关联查询。

自2011年麦肯锡年度总结报告中提出“大数据”概念以来，学术界和工业界对大数据定义一直存在争议，这些争议主要来自不同领域中大数据的特征体现[6]。目前学术界公认大数据具有“4V”特征——体量大（volume）、生成快（velocity）、多样性（variety）和密度低（value），科学大数据应用场景充分体现了这“4V”特征，并具有以下独特的性质。

（1）科学发现的准确性建立在海量实验数据的重复计算验证之上。例如，“上帝粒子”和暗物质发现的正确性经过了对数百PB量级数据的多次重复计算，多次验证重复出现同一结论时才能发布结论。

（2）短时间内科学实验会产生大量观测数据并进行流程化处理，实验数据会持续进入持久化存储设备进行长周期存储。例如，GWAC在15 s内完成40×32 MB天区图的点源检测、入库等操作，产生的所有数据将永久存储。

（3）科学现象观测的量化指标存在图像、语音、时间序列等形式，数据分布在不同国家和机构中，科学研究需要整合这些多源异构数据。例如，WDCM整合36个包括文本、网页、医疗记录在内的数据源完成知识图谱构建。

（4）科学数据来自大科学装置、互联网、国家机构等，数据与国家利益和个人隐私相关，数据共享和挖掘分析会产生更大的社会推进作用。例如，“数字丝路”（DBAR）国际科学计划涉及“一带一路”沿线65个国家共享的地理、农业、社会舆论等数据，挖掘分析这些数据可为地区、国家的决策提供重要参考，然而如何分享成果收益、保护数据隐私是该计划面临的一个重要问题。

科学大数据的这些性质对数据管理系统提出了巨大挑战。

1.2 科学大数据管理的挑战

科学大数据管理涉及数据的收集、存储、处理、分析、可视化和共享等全生命周期管理。如图 1 所示，科学应用首先从科学装置接入或从互联网采集大量异构实验或观测数据，然后经过初步过滤、转换等数据预处理操作存入持久化设备形成原始科学数据。针对具体科研目标，应用对原始数据进一步运算抽取实验特征形成特征数据。科学应用对特征数据整合挖掘分析形成科学发现量化指标，并通过可视化的方法将科学发现展现出来。最后整个流程中产生的所有数据都将存档、发布以备将来查询、验证等科研目标使用。

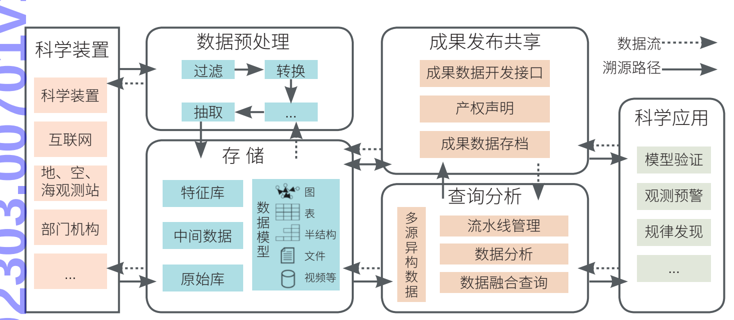


图 1　科学大数据生命周期

科学大数据管理存在常见的“4V”问题，同时也具有独特的性质，这些性质决定了科学大数据管理系统生命周期中面临 4 个方面的挑战（SPUS）。

（1）规模动态化（Scale Dynamic）。科学实验持续产生海量科学数据，并需进行长周期持久化存储。比如上文中提到的大部分科学研究项目（如 GWAC、LHC 等）每秒产生 GB 量级的观测数据，并且数据无失效期，然而科研机构却无法事先确定存储和计算资源的配置以最优地满足科学应用需求。因此，如何弹性动态地为这些数据分配存储空间和数据处理资源是科学大数据管理需要面对的一个重大挑战。

（2）流水线管理（Pipeline Management）。科学实验有严密的实验步骤，科学装置产生的海量原始科学数据会经过大量的特征提取、转换、分析等数据加工操作最终产出科研成果。以 GWAC 新星发现应用为例，原始数据进入系统以后，系统需要完成特征提取、交叉认证等严密的数据处理操作；新星预警发生后，系统需要溯源到预警产生的特征记录、天区图、镜头等并对它们进行反复确认。此外，同一个科学装置下也会出现大量类似的实验流程，因此有效地创建、执行、管理这些实验步骤和数据将极大提高科学实验的效率。

（3）统一访问（Unified Access）。大科学应用经常会对不同领域、不同机构的异构数据进行融合挖掘分析。以中国科学家发起的 DBAR 国际科学计划为例，为了给地区决策提供参考，需要获取天、空、地综合数据资源构建共享的地球大数据平台。这其中涉及卫星遥感数据、气候观测站数据、生物观测站数据以及社交网络中的舆论热点数据等异构数据的融合管理。因此，如何用统一的方式访问多源异构数据将极大地提升科学发现的价值和规模。

（4）共享管理（Sharing Management）。科学实验产生的成果数据以及中间数据通过互联开放共享以便集全世界科学家的力量进行实验验证、模型改进等后续科学研究，比如全世界物理学家通过互联网从 LHC 中获取数据进行粒子发现实验，并通过互联网共享科研成果。科学数据开放性带来的重大问题有：数据提供者与科研人员如何合理划分科研成果、数据提供者著作权认证和激励机制、共享数据的隐私保护等。如果不能妥善解决这些问题，将影响科研人员的积极性和科研生态圈的健康发展。

2 科学大数据管理系统体系架构

科学大数据管理系统主要由 4 个核心部分构成：计算和存储管理、数据流水线管理、数据融合查询管理和数据共享管理，系统体系架构如图 2 所示。计算和存储管理组件需要支持海量数据的存储和处理，并随着数据量增长动态地扩展其存储和处理能力；数据处理流程统一管理组件需要支持数据流水线的数据接入、执行、溯源和分享等一站式统一管理；数据融合管理组件需要提

供对多源异构数据的统一查询分析接口；数据共享管理组件需要规范科学发现的权益划分、数据共享的隐私保护与激励机制。



图 2　科学大数据管理系统架构

（1）计算和存储管理组件。即计算和存储资源随上层应用负载规模的变化而弹性伸缩，从而达到处理时间与资源投入的比例最优化。目前，弹性伸缩分为渐进式和定量式两种方案。渐进式伸缩方法监控上层应用对底层计算和存储资源的竞争度，动态地增加或缩减底层资源。例如，在 AWS 云平台的 E-MapReduce 集群上运行的 MapReduce 作业对资源的竞争度是集群剩余可用内存的数量，竞争度超过阈值会将新计算或存储节点纳入集群从而完成集群的自动扩容。定量式伸缩方法是通过预估目标应用的计算和存储资源需求，提前确定应用的计算和存储资源规模。与渐进式伸缩相比，定量式伸缩的反应时间较短，然而定量式伸缩方法高度依赖对目标应用的计算和对存储资源需求的准确预估[7]，如通过建立目标应用的负载模型预估系统的计算和存储资源。

（2）数据流水线管理组件。通过对数据处理流程的抽象，将数据处理过程映射为流水线中的若干逻辑处理单元，从而对数据处理过程进行规范和统一管理。通常情况下，流水线中 1 个处理单元代表 1 个函数、WebService 或 SQL 语句等，处理单元的输出可以作为其他 1 个或多个处理单元的输入；通过分支、循环等方式，这些处理单元组装在一起统一管理完成科学发现的流程。流水线管理与工作流、指令流等有相似的形式化表示，如 Pi 代数、Petri 网等[8]，通过这些流水线形式化表示，系统可在理论上保证执行过程的准确性并对异常进行捕获处理。在实际应用中，除了保证流水线的正确运行之外，流水线管理还需要解决数据接入、数据溯源、中间数据转换等核心问题，常见的流水线管理工具有 Apache Nifi、Stream Set 等。

（3）数据融合查询管理组件。即用统一的方式访问分析多源异构数据。目前数据融合主要有联邦数据库（Federate Database）、多模型数据库（Multi-model Database）、多存储数据库（Polystore Database）、数据集成（Data Integration）4 种方式[9]。联邦数据库将多个自治的异构或同构数据库中的数据透明地映射到一个全局视图中，具有自治、异源或异构、分布式的明显特征，比如在 SQL Server 2000 和 Mysql 5.0 中的 Federate 功能。多模型数据库是指一个数据库后端存储多种类型的数据，如 OrientDB、ArangoDB 等。多存储数据库架构没有统一全局视图，而是由局部视图和中间视图构成，通过统一的查询语言进行查询，典型的 Polystore 架构有 BigDAWG、Myria 等。根据数据转换的方式，数据集成可以分为在线集成和离线集成两种方式。离线集成将不同数据源中数据通过 ETL 转换，存储在全局视图数据源中进行统一管理分析，如数据仓库、数据湖泊、DataHub 等方式。在线集成通过解析查询语句将局部视图中的数据在线转换为全局视图，如 Sparksql、Impala、Presto 等[10-13]。

（4）数据共享管理组件。该组件的根本任务是疏通数据拥有者到用户之间的链路，促进数据资源在拥有者和用户之间的流通、传播与重用。目前科学数

据共享机制模式的研究主要集中在数据汇交机制、数据出版机制、数据联盟机制和服务激励机制（积分机制、在线计算服务模式）4个方面，如王晴[14]、李成赞等[15]从政策法规、技术保障、评价激励等方面对数据共享机制进行了深入分析和论证。数据共享的隐私保护技术中最具代表性的是区块链技术，如丁伟等[16]、翁健等[17]提出了基于区块链的数据共享方法，通过公私钥等非对称加密算法将数据存储在区块链上，从而更大程度上保护了用户数据的隐私，并在医疗、基因等领域进行了验证。

3 科学大数据管理系统项目进展

依托国家重点研发计划项目“科学大数据管理系统”和中国科学院“十三五”信息化建设“科学大数据工程”项目，我们与计算机领域及天文学、高能物理、微生物学等学科领域的20多家科研单位进行合作，对科学大数据管理进行了探索，研发了一套科学大数据管理系统BigSDMS（Big Scientific Data Management System）。该项目的核心内容主要包括3个部分：科学大数据管理引擎、科学大数据系统集成和科学大数据应用示范。项目研发的系统总体架构如图3所示。

3.1 科学大数据管理引擎

BigSDMS包括3类科学大数据管理引擎：大规模图数据管理、大规模半结构数据管理和大规模关系型数据管理。其中，大规模图数据库Gstore支持100亿条三元组图数据管理和秒级查询响应时间。大规模半结构化数据库Eventdb支持万亿级高能物理实验事例、EB量级数据管理能力。大规模关系型数据库AstroSever支持千亿行天文星表数据的管理，大、中、小规模数据典型操作的查询优化及满足数据处理精度与实时性的要求。这3类数据库基本满足了目前常见科学实验中大规模数据的存储、访问等管理需求。

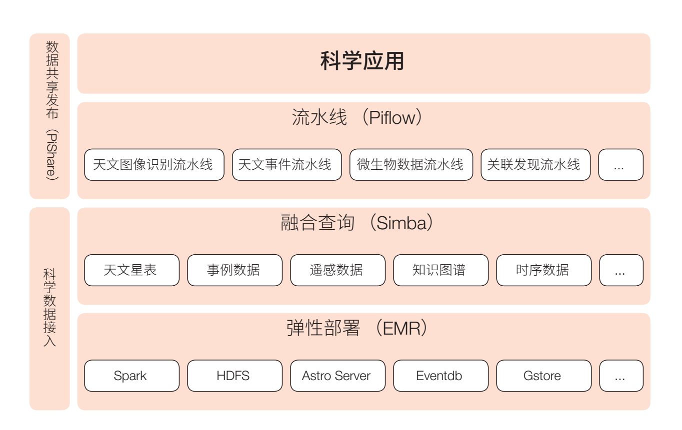


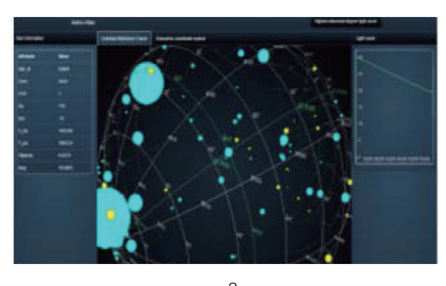
图3 BigSDMS总体架构

3.2 科学大数据系统集成

BigSDMS集成包含弹性部署（EMR）、流水线（Piflow）、融合查询（Simba）和数据共享（Pishare）4个部分。其中，EMR的弹性伸缩方案综合使用渐进式伸缩和定量式伸缩的优点：当负载模型可信度低于阈值时，采用渐进式方法进行伸缩，并根据扩容后的资源竞争修正负载模型；若负载模型可信度达到阈值后则采用定量式伸缩方法。Piflow基于Petri网，处理单元（processor）在未知状态（unknown）、活跃状态（active）、休眠状态（hibernated）3种状态之间进行转换，完成流程的执行与监控。Simba基于Sparksql，在Zeppelin可视化界面中通过SQL查询进行多种数据源的融合查询分析。Pishare基于开源区块链项目Hyperledger，在区块链上Pishare会对数据进行加密存储和产权认证，并通过积分机制（科学币）对数据提供者进行奖励以及数据市场的交易。

3.3 科学大数据应用示范

目前，基于BigSDMS，我们在天文学、高能物理、微生物学领域构建了3个应用示范：① 天文学领域使用了100亿行星表数据，定义了5个光变曲线处理流程，实现680万行星表数据插入时间少于3 s，“异常发现”时间小于1 s（图4a）；② 高能物理领域使用了BESIII产生的942.9亿条事例数据，相对于业界常用的Boss查询平均查询效率提高10倍以上（图4b）；③ 微生物学领域整合了200种微生物种菌信息，构建了5亿条规模的RDF知识图谱数据（图4c）。


a


b


c

图 4　科学大数据管理系统应用示范

（a）天文领域应用示范；（b）高能物理领域；（c）微生物学领域应用示范

4 总结与展望

随着人类对客观世界的深入认知，越来越多的社会和自然现象能够通过观测设备进行量化，这将导致科学数据的体量和类型持续增加。在数据驱动的科学发现模式下，应对科学大数据管理的 SPUS 挑战已成为眼下刻不容缓的任务。由中国科学院计算机网络信息中心牵头的国家重点研发计划“科学大数据管理系统”项目对这些问题进行了深入探索，研发了一套科学大数据管理系统 BigSDMS。未来我们还会在弹性部署、流水线、数据融合和数据发布共享 4 个方面进行更深入的探索，如竞争度的量化与预测、流水线中间数据模型设计、多查询引擎的 Polystore 方式集成、数据共享机制优化等。随着科学大数据管理技术和系统研究不断深入，科学大数据对科学发现的贡献将会越来越大！

参考文献

1 Hey T, Tansley S, Tolle K M. The Fourth Paradigm: Data-intensive Scientific Discovery. Redmond, WA: Microsoft Research, 2009.

2 黎建辉, 沈志宏, 孟小峰. 科学大数据管理：概念, 技术与系统. 计算机研究与发展, 2017, 54(2): 235-247.

3 郭华东. 大数据 大科学 大发现——大数据与科学发现国际研讨会综述. 中国科学院院刊, 2014, 29(4): 500-506.

4 杨晨, 翁祖建, 孟小峰, 等. 天文大数据挑战与实时处理技术. 计算机研究与发展, 2017, 54(2): 248-257.

5 程耀东, 张潇, 王培建, 等. 高能物理大数据挑战与海量事例特征索引技术研究. 计算机研究与发展, 2017, 54(2): 258-266.

6 Ward J S, Barker A. Undefined By Data: A Survey of Big Data Definitions. arXiv, 2013: 1309.5821v1.

7 李红臣, 史美林. 工作流模型及其形式化描述. 计算机学报, 2003, 26(11): 1456-1463.

8 Wang H, Li J, Shen Z, et al. Approximations and Bounds for (*n*, *k*) Fork-Join Queues: A Linear Transformation Approach. 2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing(CCGRID). arXiv, 2017: 1707.08860v7.

9 Lu J, Holubova I. Multi-model Data Management: What’s New and What’s Next? Extending Database Technology, 2017: 602-605.

10 Sheth A P, Larson J A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys (CSUR), 1990, 22(3): 183-236.

11 Lenzerini M. Data integration: a theoretical perspective. Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. New York: ACM, 2002: 233-246.

12 Davoudian A, Chen L, Liu M. A Survey on NoSQL Stores. ACM Computing Surveys (CSUR), 2018, 51(2): 40.

13 Gadepally V, Chen P, Duggan J, et al. The BigDAWG polystore system and architecture. IEEE High Performance Extreme Computing Computing Conference, 2016. arXiv: 1609.07548.

14 王晴. 论科学数据开放共享的运行模式、保障机制及优化策略. 国家图书馆学刊, 2014, 23(1): 3-9.

15 李成赞, 张丽丽, 侯艳飞, 等. 科学大数据开放共享：模式与机

制. 情报理论与实践, 2017, 40(11): 45-51.
16 丁伟, 王国成, 许爱东, 等. 能源区块链的关键技术及信息安全问题研究. 中国电机工程学报, 2018, 38(4): 1026-1034.
17 翁健, 李明, 张悦, 等. 一种基于区块链与代理重加密技术的可信基因检测及数据共享方法. 广东: CN108063752A, 2018-05-22.

Scientific Big Data Management Technique and System

LI Jianhui[1*]　LI Yuepeng[1,2]　WANG Huajin[1]　CHEN Mingqi[3*]

(1 Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
2 University of Chinese Academy of Sciences, Beijing 100049, China;
3 Administrative Office, Chinese Academy of Sciences, Beijing 100864, China)

Abstract As modern scientific discoveries heavily depend on the big data management, it is an urgent task to research how to manage scientific big data efficiently. In this paper, we first introduce the application scenes and requirement of scientific big data. Then we summarize four challenges in the management of scientific big data (SPUS): Scale dynamic, Pipeline management, Unified access, and Sharing management. After that, we present the proposed scientific big data management system which consists of four components: computing & storage management, data processing management, data fusion management, and data sharing management. Moreover, we specify the key techniques in the proposed system. At last, we introduce the ongoing Big Scientific Data Management System (BigSDMS) program, which is a national key research and development program.

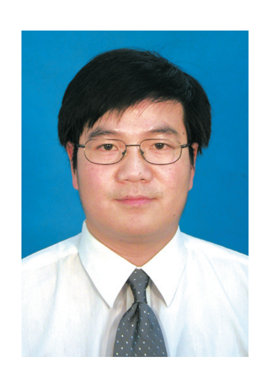
Keywords scientific big data, integrate query, pipeline, data sharing, elastic expansion



黎建辉 中国科学院计算机网络信息中心研究员，博士，博士生导师；国际科技数据委员会（CODATA）执委，大数据应用服务技术北京市工程实验室主任。长期致力于推动科学数据开放、共享与应用服务，负责中国科学院科学数据云服务体系构建、云服务平台研发以及数据密集型科研应用创新工作。当前主要从事大数据资源开放共享、大数据管理技术、大数据计算与分析技术等方面研究工作。E-mail: lijh@cnic.cn

LI Jianhui Professor in Computer Network Information Center, Chinese Academy of Sciences (CAS). He is also the Executive Member at CODATA, and Director of Beijing Engineering Laboratory for Big Data Application Service Technologies. Prof. Li has long been dedicated to promoting data openness, sharing, and application. His major tasks include: to construct the CAS Data Cloud Service System, to develop cloud service platforms, and to innovate data-intensive scientific applications. He is currently engaged in technologies concerning big data sharing, curation, computing, and analysis.
E-mail: lijh@cnic.cn

*Corresponding author



陈明奇　中国科学院网信办副主任、办公厅网信处处长，博士。研究方向：信息化战略、科研信息化战略、网络与信息安全等。自2007年起负责中国科学院信息化建设规划及实施工作，参与国家及中科院的多项信息化工程、网络安全工程建设，参与中国科学院院士咨询项目"国家科研信息化战略研究"（2014—2015年）、"前沿与交叉学科科研信息化发展战略研究"（2016—2017年），组织编撰《中国科研信息化蓝皮书》《中国科学院信息化发展报告》《中国科学院信息化评估报告》等系列报告。发表学术文章40余篇，出版译著1本。E-mail: mqchen@cashq.ac.cn

CHEN Mingqi　Director of Informatization Division, General Office, Chinese Academy of Sciences (CAS). He received his doctorate degree in Information and Signal Processing from Beijing University of Posts and Telecommunications in 2000. From 2007, he has been responsible for planning and implementation of CAS information system construction, as well as daily operation and maintenance of information system. He coordinates and supports the acquisition, development, and provision of state-of-the-art information infrastructure resources, tools, services, and information application systems essential to CAS. He is the member of the Editorial Board of "China's Blue Book on Scientific Research Informatization". His main research areas include information technology, scientific research information, network and information security, signal and information processing, etc. E-mail: mqchen@cashq.ac.cn

■责任编辑：岳凌生